COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13294
© 2023 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).
ISSN: 1551-6709 online
DOI: 10.1111/cogs.13294

Prediction From Minimal Experience: How People Predict the Duration of an Ongoing Epidemic

Yi-Long Lu,[a,#] Yang-Fan Lu,[b,c,#] Zhuo Rachel Han,[d] Shaozheng Qin,[e,f] Xin Zhang,[a] Li Yi,[a,g] Hang Zhang[a,c,f,g]

[a]*School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University*
[b]*Academy for Advanced Interdisciplinary Studies, Peking University*
[c]*Peking-Tsinghua Center for Life Sciences, Peking University*
[d]*Faculty of Psychology, Beijing Normal University*
[e]*State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University*
[f]*Chinese Institute for Brain Research, Beijing*
[g]*PKU-IDG/McGovern Institute for Brain Research, Peking University*

Received 22 June 2022; received in revised form 7 April 2023; accepted 18 April 2023

Abstract

People are known for good predictions in domains they have rich experience with, such as everyday statistics and intuitive physics. But how well can they predict for problems they lack experience with, such as the duration of an ongoing epidemic caused by a new virus? Amid the first wave of COVID-19 in China, we conducted an online diary study, asking each of over 400 participants to predict the remaining duration of the epidemic, once per day for 14 days. Participants' predictions reflected a reasonable use of publicly available information but were meanwhile biased, subject to the influence of negative affect and future time perspectives. Computational modeling revealed that participants neither relied on prior distributions of epidemic durations as in inferring everyday statistics, nor on mechanistic simulations of epidemic dynamics as in computing intuitive physics. Instead, with minimal experience, participants' predictions were best explained by similarity-based generalization of the temporal pattern of epidemic statistics. In two control experiments, we further confirmed that such cognitive algorithm is not specific to the epidemic scenario and that minimal and rich experience do lead to different

[#]co-first authors.
Correspondence should be sent to Hang Zhang and Li Yi, School of Psychological and Cognitive Sciences, Peking University, Beijing, China. 100871. E-mail: hang.zhang@pku.edu.cn and yilipku@pku.edu.cn

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

prediction behaviors for the same observations. We conclude that people generalize patterns in recent history to predict the future under minimal experience.

Keywords: Prediction; Bayesian inference; Function learning; Crowd wisdom; Folk epidemiology; COVID-19

1. Introduction

Predicting the future is an ability that everyone dreams of and, in some sense, we do have. For example, when people are asked to predict everyday statistics such as the expected life span of a man of 90 years old or the total box office of a movie that has grossed 10 million dollars, their predictions are strikingly consistent with the real-world statistics, at least on the group level (Griffiths & Tenenbaum, 2006, 2011; Mozer, Pashler, & Homaei, 2008). It is as if they use Bayesian inference to solve the prediction problem (Griffiths & Tenenbaum, 2006), combining the current observation of a quantity (e.g., 90 years old) with their prior knowledge of its probability distribution (e.g., the life-span distribution of the population). A second class of prediction problems in the literature of human cognition[1] is known as intuitive physics (Kubricht, Holyoak, & Lu, 2017). It is found that people can accurately predict the future status of physical events that evolve with time (Hamrick, Battaglia, Griffiths, & Tenenbaum, 2016; Kubricht et al., 2017), such as the trajectory of a freely falling object, as if they could simulate ongoing dynamics based on appropriate physical laws (Battaglia, Hamrick, & Tenenbaum, 2013; Hamrick et al., 2016; Smith, Battaglia, & Vul, 2018). Everyday statistics and intuitive physics represent two classes of prediction problems associated with different levels of theories (see Fig. 1): The former relies on *descriptive theories* that directly char-

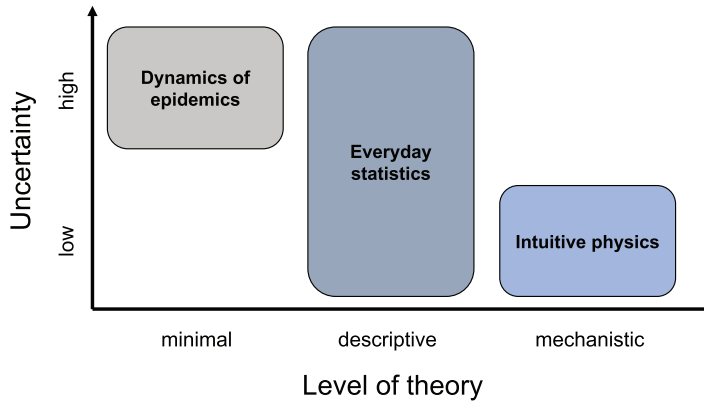

Fig. 1. Three classes of representative prediction problems in human cognition.

Note. The three classes of problems differ in two dimensions: whether people have mechanistic, descriptive, or minimal theories about the problem ("level of theory"), and whether the uncertainty associated with the problem is low or high ("uncertainty").

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

acterize the distribution of similar quantities (e.g., life spans of other individuals), while the latter on *mechanistic theories* of how the quantity in question (e.g., trajectory) is controlled by other quantities (e.g., velocity and acceleration). Despite these differences, both belong to problems people have rich experience with.

Here, we focus on a third class of prediction problems: those that people have minimal experience with, such as the dynamics of an ongoing epidemic caused by a new virus. The outbreak of the COVID-19 pandemic has reminded us of the real-life importance of such predictions. Scientists all around the world have been developing mathematical models to predict the progress of the pandemic (Dehning et al., 2020; Maier & Brockmann, 2020), which can help governments to respond proactively and better balance the costs and benefits of intervention policies, such as school closures and traveling restrictions (Flaxman et al., 2020; Hsiang et al., 2020). Common people's predictions, though less publicized in the media, may also significantly change the trajectory of the pandemic, through both the public's opinions (Burstein, 2003) and individuals' conformity with the recommended protective behaviors, such as face-mask wearing and social distancing (Rubin, Amlot, Page, & Wessely, 2009). Previous studies on folk epidemiology only tested people's medical knowledge about infectious diseases, such as the estimation of disease prevalence (Kalichman & Cain, 2005) or the understanding of disease causality (Au et al., 2008; Legare, Evans, Rosengren, & Harris, 2012; Sigelman & Glaser, 2019). It is largely unknown how people may use available information to predict the dynamics of an ongoing epidemic.

Due to its distinctive characteristics, the prediction of epidemic dynamics likely requires different cognitive computations from those of everyday statistics and intuitive physics. On one hand, because pandemics or even large-scale epidemics were not common in one's lifetime, most people may not be able to base their predictions on experienced distributions of similar quantities, as they do for everyday statistics. On the other hand, predicting the spread of a new infectious disease can be much more difficult than predicting physical events, despite that both involve dynamic processes. This difficulty arises not only from people's limited experience with the disease, but also from the higher uncertainty associated with disease transmission than with physical events. Both environmental factors (Merow & Urban, 2020) and human behaviors (Flaxman et al., 2020; Hsiang et al., 2020) add uncertainty to the spread of the disease. Even for mathematical models that are designed to forecast the trend of COVID-19 (Bertozzi, Franco, Mohler, Short, & Sledge, 2020; Dehning et al., 2020; Estrada, 2020; Maier & Brockmann, 2020), the complexity of influencing factors remains a challenge. It is thus reasonable if people use more heuristic computations instead of computing epidemic dynamics as they do for intuitive physics.

The unfortunate outbreak of COVID-19 provides us a real-life situation to test human predictions of epidemic dynamics. During February and March 2020, when COVID-19 had not developed into a global pandemic but was mainly reported as an epidemic outbreak in China, we recruited over 400 Chinese adults online to participate in a daily survey study, where each participant reported their predictions for the remaining duration of the epidemic (i.e., until when the number of daily new cases in China comes back to zero), once per day for 14 days. They were also assessed on a range of psychological traits and states. Different from many laboratory prediction tasks (Quiroga, Schulz, Speekenbrink, & Harvey, 2018), our survey did

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

not supply any epidemic statistics or suggest which information should be relevant. That is, participants faced a genuine real-world problem and were free to use whatever information they possessed to make predictions.

To understand the cognitive computations underlying participants' predictions, we constructed four computational models and compared their goodness-of-fit to human data. The first model implements the Bayesian inference used for everyday statistics, combining the number of elapsed days with a prior distribution of epidemic durations (Griffiths & Tenenbaum, 2006). The second model is analogous to the dynamics computation in intuitive physics, which mimics the modeling of dynamics in a state-of-the-art mathematical model of epidemics (Maier & Brockmann, 2020). Alternative to building an internal model to simulate the dynamics of epidemics, people may adopt a heuristic approach that is more economical in computation, focusing on the historical trend of the epidemic statistic in question to predict its trend in the future. Learning and generalization like this is known as function learning, through which people can learn from limited instances and generalize the observed input-output mapping to unlearned inputs (DeLosh, Busemeyer, & McDaniel, 1997; Lucas, Griffiths, Williams, & Kalish, 2015; Schulz, Tenenbaum, Duvenaud, Speekenbrink, & Gershman, 2017). The third and fourth models, respectively, implement rule-based and similarity-based (also known as parametric and nonparametric) function learning algorithms. For rule-based function learning, the observer expects the observed epidemic statistic to change with time following a specific bell-shaped functional form and predicts the future by estimating the parameter(s) of this function. For similarity-based function learning, the observer has no assumption about the functional form of the trend and predicts the future as the weighted average of previous observations, with observations from farther past carrying smaller weights.

We also performed two control experiments using a cover story that is irrelevant to disease, to exclude the possibility that our findings in the survey study were specific to the epidemic scenario or the particular trend of observations, and to further compare human prediction behaviors under rich and minimal experiences.

2. Methods

2.1. Participants and survey administration

The study was part of a collaborative project of multiple labs in Peking University and Beijing Normal University. It was conducted during the first wave of COVID-19 in China, when the cases outside China were still few (Fu et al., 2021; Li, Lai, Gao, & Shi, 2021). Via the online survey software Qualtrics, we administered four rounds of surveys, which started on February 6, February 10, February 17, and February 29 in 2020. Participants were Chinese-speaking adults who were located in China, recruited through the I-love-experiments platform (http://aishiyan.bnu.edu.cn/), an online participant pool maintained by Beijing Normal University. The study had been approved by the Institutional Review Boards of School of Psychological and Cognitive Sciences at Peking University and Faculty of Psychology at

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Beijing Normal University. All participants provided informed consent online. Participants were compensated for their time, receiving 10 RMB for a baseline assessment and 2 RMB for each valid daily entry.

Each round of surveys consisted of a 30-min baseline assessment performed on the initial test and a 5-min daily survey (i.e., diary) administered on 14 consecutive days. Each participant only took one round of surveys. During the baseline assessment, demographic information was collected, all items in the daily survey were administered, and a series of personal traits and emotional states were assessed. After the baseline assessment, participants who volunteered to participate in the following 14-day daily surveys and also met the inclusion criteria (passing the attention check and rationality check, see *Baseline Assessment* and *Daily Measures* for details) were registered into the daily survey program. On each day, participants must fill in the survey between 6:00 p.m. and 11:59 p.m., so that the epidemic statistics (new cases, recoveries, and deaths, etc.) for the previous day had been reported in the media. To help participants complete the survey on time, experimenters provided daily reminders for each participant. We also controlled the settings of Qualtrics to ensure that the survey was accessible only within the required time range.

We intended to maximize the sample size within our budget limit and recruited every adult volunteer who responded to our advertisement to participate in the baseline assessment. Among the 1039 participants who completed the baseline assessment, 223 participants failed the attention check or rationality check. In the remaining 816 participants, 420 participants volunteered to participate in the 14-day daily surveys. They missed 113 and completed 5767 daily survey entries, among which 228 entries (3.95%) were excluded for failing the rationality check or taking more than 3 h to complete. Two participants failed the rationality check for all their entries. The final sample for the current analysis thus included 5539 daily surveys from 418 participants (aged 18–50, mean = 24.1, 80.5% female, see Fig. S3 for the number of participants on each calendar date).

2.2. Baseline assessment

Baseline assessment includes demographic information and 11 self-report psychological scales. Among them, three scales measure the levels of negative affect, including depression, anxiety, and stress: Depression Anxiety Stress Scales-21 (DASS-21; Taouk, Lovibond, & Laube, 2001), Perceived Stress Scale (PSS; Chu & Kao, 2005; Cohen, Kamarck, & Mermelstein, 1983), and Trait Anxiety Inventory (T-AI; Spielberger et al., 1983; Wang, Wang, & Ma, 1999; Zheng et al., 1993). The Future Time Perspective Scale (FTP scale; Carstensen & Lang, 1996) measures the perspective toward the future. Please see the Supplement for more details and a full list of all the scales. Scales were administered in a simplified Chinese version. Three additional instructed choices (e.g., "Please select very unlikely for the current question") were implanted among the scales to serve as attention checks.

2.3. Daily measures: Predicting the duration of the epidemic

In February 2020, COVID-19 was known as an ongoing epidemic in China instead of a pandemic. At that time, many Chinese people, including the authors, had expected it to end

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

soon, similar to the SARS epidemic in 2003. Participants were required to predict the number of days it would take before the daily new cases of COVID-19 in China went to zero. In particular, four questions were presented in the following order: (1) "After how many days **at least** do you think there will be no more daily new cases in our country?" (i.e., the minimum duration in their own belief); (2) "After how many days **at most** do you think there will be no more daily new cases in our country?" (i.e., the maximum duration in their own belief); (3) "What do you think this duration (in days) is **most likely** to be?" (i.e., the most possible duration in their own belief); and (4) "What do you think **most people's prediction** for this duration will be?" (i.e., the most likely duration in others' belief). Participants answered each question by typing in a non-negative number for days. We used this free inputting format instead of forced choices or the Likert scale to avoid any anchoring effects.

Answers to the first three questions were used as a rationality check that requires the minimum $\leq$ the most possible $\leq$ the maximum. Participants' predictions were added by one and then *log-transformed* for further analysis. Unless otherwise noted, our further analysis of the duration prediction focused on the "most likely" duration.

2.4. COVID-19 epidemic statistics

The epidemic statistics of COVID-19 used in our analyses (daily new cases, cumulative cases, etc.) came from the National Health Commission (NHC) of the People's Republic of China (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml).

We did not provide these COVID-19 data to participants in our survey but expected participants to have access to the information themselves from everyday news. NHC published daily COVID-19 reports with a delay of 1 day (e.g., NHC published February 9's data on the morning of February 10). Therefore, the latest COVID-19 data participants could access when filling in the daily surveys were from the previous day.

The daily new cases could be divided into local new cases and imported new cases. We used local new cases alone to define the ground truth of the zero-case day. In regression and modeling analyses, we used the total new cases (i.e., local + imported).

2.5. Modeling the prediction of epidemic duration

To understand the cognitive process behind participants' predictions of the epidemic duration, we formulized the three types of computations described in the Introduction into four computational models—the duration-prior model, the epidemiologist model, the rule-based function learning model, and the similarity-based function learning model—and compared their goodness-of-fit to human behavior (see Fig. 3A for an illustration). Below, we specify the assumptions of each of these models from three aspects: the observer's belief of how the observed epidemic statistics are generated ("generative model"), how the observer infers the latent parameters or variables in the generative model ("inference"), and how the observer uses the fitted generative model to predict the future ("prediction"). Please also see Table S2 for a summary of the free parameters for each model.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.5.1. Duration-prior model

In the duration-prior model, we consider an observer with some prior beliefs about the distribution of the epidemic's durations. To predict the future, the observer combines the prior distribution with the actual duration that has elapsed since the epidemic onset following Bayesian inference (Griffiths & Tenenbaum, 2006, 2011). Griffiths and Tenenbaum (2011) used the Erlang distribution to approximate human prior for the duration of daily events, because it "provides a simple way to summarize many of the kinds of distributions that might be encountered across temporal domains" (Griffiths & Tenenbaum, 2011, p. 5). In a similar rational, here we chose the gamma distribution—a generalized form of the Erlang distribution—as the prior belief in the duration-prior model. Following previous studies on the prediction of everyday statistics (Griffiths & Tenenbaum, 2006, 2011), we fit the parameters of the prior distribution as free parameters.

Generative model: We assume that the observer's prior belief for the epidemic's total duration t_{total} follows a Gamma distribution:

$$t_{\text{total}} \sim \text{Gamma}\,(a,\ b)\,, \tag{1}$$

where $a > 1$ controls the shape, and $b > 0$ controls the scale. After a specific duration t_{elapsed}, when the duration has not ended, the likelihood function for t_{total} equals $p\,(t_{\text{total}} \geq t_{\text{elapsed}} | t_{\text{total}}) = 1$ for all $t_{\text{total}} \geq t_{\text{elapsed}}$, and 0 otherwise.

Inference: Given the duration's prior and the likelihood function, the poster estimation of the epidemic duration t_{total} can be inferred:

$$p\left(t_{\text{total}} | t_{\text{total}} \geq t_{\text{elapsed}}\right) = \frac{p\left(t_{\text{total}} \geq t_{\text{elapsed}} | t_{\text{total}}\right) p\,(t_{\text{total}})}{\int p\left(t_{\text{total}} \geq t_{\text{elapsed}} | t_{\text{total}}\right) p\,(t_{\text{total}})\,\mathrm{d}t_{\text{total}}}, \tag{2}$$

which yields

$$p\left(t_{\text{total}} | t_{\text{total}} \geq t_{\text{elapsed}}\right) = \frac{f\,(t_{\text{total}}, a, b)}{1 - F\left(t_{\text{elapsed}}, a, b\right)}, \tag{3}$$

where $f(t, a, b)$ and $F(t, a, b)$, respectively, denote the probability density function and cumulative distribution function of Gamma(a, b).

Prediction: Following Griffiths and Tenenbaum (2006, 2011), the Bayesian observer calculates the posterior median $\hat{t}_{\text{total}}$ of the epidemic duration t_{total} given the number of elapsed days t_{elapsed}. On the test day t_{test}, we assume that the elapsed days according to the observer's memory is $t_{\text{elapsed}} = t_{\text{test}} + t_0$, where t_0 is a free parameter reflecting the observer's estimation of the number of epidemic days before the survey. We assume that the observer's estimation of the remaining duration of the epidemic, D_{end}, is further contaminated by an additional Gaussian noise $\epsilon_D \sim \text{N}(0,\ \sigma^2_{\text{est}})$ on the logarithmic scale:

$$\ln\,(D_{\text{end}}) = \ln\left(\hat{t}_{\text{total}} - t_{\text{test}} - t_0 + 1\right) + \epsilon_D, \tag{4}$$

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In total, the duration-prior model has four free parameters: Gamma distribution prior parameters a and b, memory reference point t_0, and estimation noise σ^2_{est}.

2.5.2. Epidemiologist model

Consider an observer who is armed with epidemiological knowledge and uses this knowledge to model the dynamics of disease transmission. To implement this "epidemiologist" observer, we adapted a Susceptible-Infected-Recovered (SIR) model recently published in *Science* (the SIR-X model; Maier & Brockmann, 2020), which successfully captures the effects of the Chinese government's containment and quarantine procedures. By modeling not only the effects of the containment procedure influencing almost everyone but also those of the quarantine procedure removing only the symptomatic infected individuals, the SIR-X model can predict the subexponential growth of the infected population observed in China that violates the predictions of other state-of-the-art SIR models (e.g., Dehning et al., 2020).

Generative model: The observer believes that from day t to day $t+1$, the number of susceptible individuals (S), the number of infected individuals who have not been identified and thus are not quarantined (I), the number of individuals who are permanently removed from the susceptible or unquarantined infected due to recovery or containment procedures (R), and the cumulative number of quarantined infected individuals (X), respectively, undergo changes:

$$S_{t+1} - S_t = -\beta \frac{I_t S_t}{N} - \kappa_0 S_t, \tag{5}$$

$$I_{t+1} - I_t = \beta \frac{I_t S_t}{N} - (\gamma + \kappa_0 + \kappa) I_t, \tag{6}$$

$$R_{t+1} - R_t = \gamma I_t + \kappa_0 S_t, \tag{7}$$

$$X_{t+1} - X_t = (\kappa_0 + \kappa)\ I_t. \tag{8}$$

Here, $N = S_t + I_t + R_t + X_t$ is the total population, β is the transmission rate that controls the spread of the disease from the infected to the susceptible, γ is the recovery rate of the infected, κ_0 is the general containment rate that affects both the susceptible and infected populations, and κ captures the effects of the policies such as quarantine measures that only affect the infected.

Inference: Among the four series of variables (S, I, R, and X), only X is observable through the reported cumulative cases. Following the original paper of this SIR-X model (Maier & Brockmann, 2020), we set $\beta = 0.775$ and $\gamma = 0.125$. We assume that the observer infers the values of the epidemic parameters κ_0, κ and the initial status I_0 from the time series of daily epidemic statistics (cumulative cases, or equivalently, daily new cases) so that the mean squared errors between the observed and predicted time series of X are minimized.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Prediction: On the test day t_{test}, the observer can compute a prediction of epidemic statistics (*S, I, R, X*) for any future day $\tau > t_{\text{test}}$ based on Eqs. 5–8, with the difference equations starting from the first day when daily epidemic statistics were publicly available (January 21, 2020) and the epidemic parameters estimated using the statistics up to day $t_{\text{test}} - 1$.

The ending day t_{end} of the epidemic is defined as the first day when the expected number of new cases is less than a logarithmic threshold θ. That is, $t_{\text{end}} \geq t_{\text{test}}$ and satisfies:

$$\ln\left(X_t - X_{t-1}\right) \geq \theta,\ \forall t \in \{1, 2, \ldots, t_{\text{end}} - 1\}; \tag{9}$$

$$\ln\left(X_{t_{\text{end}}} - X_{t_{\text{end}}-1}\right) < \theta. \tag{10}$$

We assume that the observer's estimation of the remaining duration of the epidemic, D_{end}, is further contaminated by an additional Gaussian noise $\epsilon_D \sim \text{N}(0,\ \sigma^2_{\text{est}})$ on the logarithmic scale:

$$\ln(D_{\text{end}}) \ = \ \ln(t_{\text{end}} - t_{\text{test}} + 1) + \epsilon_D. \tag{11}$$

In total, the epidemiologist model has two free parameters to be estimated from participants' predictions: zero-case threshold θ and estimation noise σ^2_{est}.

2.5.3. Similarity-based function learning model

Different from an epidemiologist observer who bases on an internal model to compute the dynamics of disease transmission, people may have limited knowledge to guide their prediction, except for the epidemic statistics reported in the news. A more knowledge-free way to predict the future is to use similarity-based function learning: The closer the two dates are, the more similar their number of new cases will be. In other words, the case number at a future time point is expected to depend on past observations of case numbers, with the impact of past observations decaying with time. Formally, the similarity-based function learning can be implemented by the Gaussian process regression (Rasmussen & Williams, 2005) with a radial basis function (RBF) kernel (Stojic, Schulz, Analytis, & Speekenbrink, 2020). With an RBF kernel, the similarity-based function learning model hypothesizes that with no more cases observed, the cases in the future would drop to a certain level and the epidemic would finally stop (i.e., the prior mean is 0).

Generative model: From the perspective of function learning, to predict the future growth of COVID-19 is to learn from the observed data patterns of daily new cases and generalize the pattern to future time points. Consider an arbitrary series of days $\boldsymbol{T} = [t_0, t_1, \ldots, t_n]^\top$, where t_i and t_{i+1} are not necessarily adjacent days. The observer assumes that the logarithm of the true number of daily new cases on these days, $f(\boldsymbol{T})$, is generated from a Gaussian process, that is, a multivariate Gaussian prior distribution with zero mean and covariance $k(\boldsymbol{T},\ \boldsymbol{T}^\top)$:

$$f\left(\boldsymbol{T}\right) \sim N\left(0,\ \begin{bmatrix} k\left(t_0, t_0\right) & \cdots & k\left(t_0, t_n\right) \\ \vdots & \ddots & \vdots \\ k\left(t_n, t_0\right) & \cdots & k\left(t_n, t_n\right) \end{bmatrix}\right), \tag{12}$$

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $k(t_i, t_j)$ is the kernel function that characterizes the covariance between 2 days t_i and t_j. Here, in the similarity-based model, we have:

$$k\left(t_i,\ t_j\right) = \exp\left(-\frac{\left|t_i - t_j\right|^2}{2\lambda^2}\right), \tag{13}$$

where λ is the length-scale that controls how the similarity (or correlation) decays with the distance between inputs (i.e., the time interval between dates). The observer also assumes that the observed number of daily new cases on any specific day, Y_t, is contaminated by an additional Gaussian noise $\epsilon_t \sim N(0,\ \sigma^2_{\text{obs}})$:

$$Y_t = f\left(t\right) + \epsilon_t. \tag{14}$$

Practically, Y_t is the daily new cases reported in the news (equivalent to the $X_t - X_{t-1}$ above), while $f(t)$ is the true daily new cases that are unobservable.

Inference and prediction: On the test day t_{test}, the observer has access to the reported daily new cases up to day $t_{\text{test}} - 1$. Denote the series of days from day 0 to day $t_{\text{test}} - 1$ by $\boldsymbol{T}_{\text{test}} = [0, 1, \ldots, t_{\text{test}} - 1]^\top$ and its reported daily new cases by $\boldsymbol{Y}_{\text{test}} = [Y_0, Y_1, \ldots, Y_{t_{\text{test}}-1}]^\top$. We assume that the observer would predict the daily new cases $f(\tau)$ on any specific future day $\tau > t_{\text{test}}$ using Bayesian inference. In terms of Gaussian process regression, this is equivalent to first computing the joint distribution

$$\begin{bmatrix} \boldsymbol{Y}_{\text{test}} \\ f\left(\tau\right) \end{bmatrix} \sim N\left(0,\ \begin{bmatrix} k\left(\boldsymbol{T}_{\text{test}}, \boldsymbol{T}_{\text{test}}\right) + \sigma^2_{\text{obs}}\boldsymbol{I} & k\left(\boldsymbol{T}_{\text{test}}, \tau\right) \\ k\left(\tau, \boldsymbol{T}_{\text{test}}\right) & k\left(\tau, \tau\right) \end{bmatrix}\right), \tag{15}$$

and then marginalizing off $\boldsymbol{Y}_{test}$ to obtain the posterior $f(\tau)|\boldsymbol{Y}_{\text{test}}, \boldsymbol{T}_{\text{test}}$, which follows a Gaussian distribution.

The report of epidemic duration is elicited from the posterior mean function of daily new cases (i.e., the mean of $f(\tau)|\boldsymbol{Y}_{\text{test}}, \boldsymbol{T}_{\text{test}}$ as a function of τ) in the same way as in the epidemiologist model.

In total, the similarity-based function learning model has four free parameters: length-scale λ, observation noise σ^2_{obs}, zero-case threshold θ, and estimation noise σ^2_{est}.

2.5.4. Rule-based function learning model

As an alternative type of function learning, an observer who follows rule-based function learning would fit a specific parametric family to the observed time series of daily new cases and use the fitted function to predict the future. In particular, the observer believes that the number of daily new cases first increases to a "turning point" and then gradually decreases to zero. Similar to the similarity-based function learning model, the observer's inference and prediction of the rule-based model is implemented by the Gaussian process regression, though with a different kernel.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Generative model: The observer's prior belief is formulized as a quadratic function on the logarithmic scale and implemented by a Gaussian process with a quadratic kernel:

$$k\left(\boldsymbol{T}, \boldsymbol{T}'\right) = \left(\gamma \boldsymbol{T}^{\top} \boldsymbol{T}' + c_0\right)^2, \tag{16}$$

where γ is the scale parameter and c_0 is the offset parameter. A quadratic function on the logarithmic scale corresponds to a bell-shaped curve on the linear scale, which agrees with the dynamics that emerge from SIR models.

Inference and prediction: The inference and prediction of the rule-based model follow the Gaussian process described above for the similarity-based model.

In total, the rule-based function learning model has five free parameters: scale parameter γ, offset parameter c_0, observation noise σ^2_{obs}, zero-case threshold θ, and estimation noise σ^2_{est}.

2.6. Model fitting and comparison

We averaged the duration predictions across participants separately for each test day and fit each model to these aggregated duration predictions using maximum likelihood estimates. The Akaike Information Criterion (AIC; Akaike, 1974) was used to compare the goodness-of-fit of different models. For each model, ΔAIC was calculated as the difference between the AIC of the model and the minimum AIC among all four models. AIC weight, a measure that reflects the relative likelihood of a specific model to outperform the other models, was computed for each model as $\frac{exp(-\frac{1}{2}\Delta AIC_{the})}{\sum_j exp(-\frac{1}{2}\Delta AIC_j)}$, where ΔAIC_{the} denotes the ΔAIC of the model and ΔAIC_j denotes the ΔAIC of model j, with j enumerating all the four models (Burnham & Anderson, 2002).

As a second metric for model comparison, we also performed a five-fold cross-validation as follows. We split the 40-day daily surveys into five subsets of 8 days. The first and last days were omitted for small sample sizes, so that the first and last subsets only included 7 days. For each model, each time one of the five subsets served as the test set and the remaining four subsets as the training set. The summed log-likelihood across five test sets was used as the goodness-of-fit metric for model comparison. Model fitting was implemented in Python 3.8 with the constrNMPy (v0.1) package. The optimization process was repeated for 1000 times with random starting points to ensure the convergence of the negative log-likelihood to the global minimum.

2.7. Statistical analysis

2.7.1. Linear mixed-effects models

We performed a series of linear mixed-effects models (LMMs) to see (1) which epidemic statistic best predicted participants' duration predictions; (2) how a participant's DASS and FTP scores may modulate the participant's duration prediction; and (3) the participant's duration prediction "pessimistic-me bias" (defined in Results). Please see the Supplement for a full description of the LMMs. All data were transformed to z-score for regression. All LMM analyses were implemented in python3.8 with the *statsmodels* (v0.13.1) package.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.7.2. Correction for multiple comparisons

To avoid the inflation of Type I errors due to multiple comparisons, we performed the false discovery rate (FDR) correction in our analysis, using the *fdrcorrection* function in the Python package *statsmodels*.

3. Results

3.1. Behavioral signatures in human predictions of epidemic durations

The daily new cases included local new cases and imported new cases. Given that the latter was unexpected at the beginning of our study (February 6, 2020), we defined the first day when the local new cases became zero as the ground truth of the zero-case day, which was March 18, 2020.

We first examined all 418 participants as a collective mind, averaging their duration predictions separately for each test day (diary date) and contrasting them with the ground truth (Fig. 2A). Across the test period of 40 days, the aggregated predictions for epidemic durations ranged from 24 to 42 days, which on average were significantly greater than the ground-truth durations (Wilcoxon signed-rank test, $Z = 5.22$, $p < .001$) by 12.5 days.

On one hand, participants' predictions were reasonable: The predicted durations had an overall decreasing trend as days elapsed (linear regression, slope $\beta = -0.247$, 95% CI $[-0.270, -0.223]$, $p < .001$). On the other hand, the decrease was not fast enough to converge to a specific date. In contrast, the SIR-based models used by epidemiologists (e.g., the SIR-X model) would predict the duration to converge as the infective population decreased to zero or a stable endemic level (Allen, 1994). The zero-case day implied from participants' duration predictions had a tendency to extend to a farther and farther future (linear regression, slope $\beta = 0.230$, 95% CI [0.208, 0.253], $p < .001$), which we call the "postpone-again bias." This bias existed even during the first 27 test days when no imported cases had been reported (linear regression, slope $\beta = 0.348$, CI [0.315, 0.380], $p < .001$). That is, the bias was not simply because participants changed their minds after knowing of the possibility of imported cases.

Because no epidemic statistics were presented in our daily survey, participants could only use the epidemic statistics (up to yesterday) that were available in the news, which included daily new cases (local + imported), cumulative cases, total active cases, daily deaths, cumulative deaths, daily recoveries, and cumulative recoveries. Among these statistics, if we only consider one single statistic from yesterday, it would be the daily new cases (see Fig. S1) that correlated most with the duration predictions of an ideal observer following a state-of-the-art mathematical model of epidemics (Maier & Brockmann, 2020). Did participants' predictions reflect a similar highest contribution from the daily new cases?

Given the high correlations between the statistics (see Fig. S4), we entered different epidemic statistics from yesterday into separate LMMs (see Supplemental Methods) as predictors for participants' epidemic durations. We then compared the resulting goodness-of-fit of different models to see which statistic contributed most. To rule out the possibility that the

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

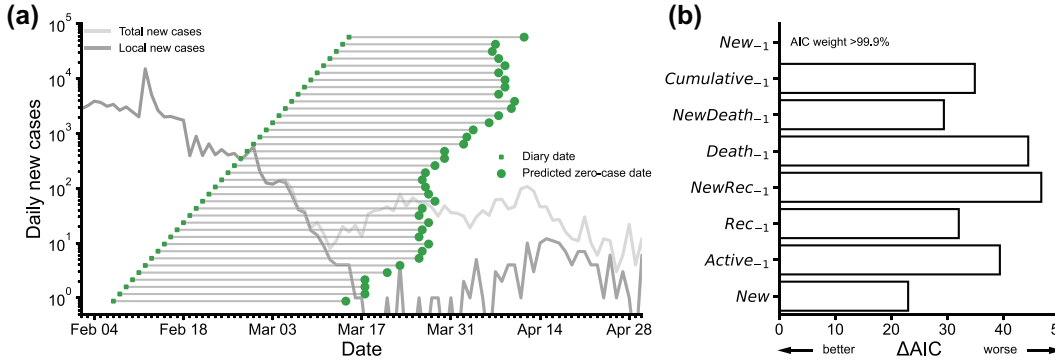

Fig. 2. Behavioral signatures in human predictions of epidemic durations.

Note. Panel A: Daily new cases and predicted epidemic durations for each day. The test period of our diary program ranged from February 6, 2020 to March 16, 2020 (Data from the first and last days were omitted due to small sample sizes). Gray curves denote the daily new cases for each date in Mainland China (reported by the National Health Commission of China), with darker gray for local new cases and lighter gray for total new cases (i.e., local + imported), both shown in logarithmic scale. Each horizontal gray line denotes the predicted "most likely" duration on a specific test day (i.e., diary date), averaged across participants. The green rectangle at its left end denotes the diary date. The green circle at its right end denotes the implied zero-case date (i.e., diary date + predicted duration). Throughout the test period, participants' duration predictions had a "postpone-again bias": the implied zero-case date extended to farther and farther future instead of converging to a specific date. Panel B: Comparison of different epidemic statistics as linear predictors for participant's duration predictions. New_{-1}, $Cumulated_{-1}$, $NewDeath_{-1}$, $Death_{-1}$, $NewRec_{-1}$, Rec_{-1}, and $Active_{-1}$, respectively, denote yesterday's new cases, cumulative cases, daily deaths, daily recoveries, cumulative recoveries, and total active cases, which were the latest epidemic statistics participants could access on the test day. *New* denotes today's new cases, which was inaccessible and served as a control test. Each of these variables was entered into a separate linear mixed-effects model as predictors for participants' duration predictions. Smaller ΔAIC indicates better fits. Yesterday's new cases best predicted participants' duration predictions.

observed influence of daily new cases on participants' duration predictions might be an artifact reflecting their common decreasing trend, we also included a control model with today's new cases (i.e., information unavailable at test) as the predictor. According to the metric of goodness-of-fit, ΔAIC (the lower the better), the regression model that best fit participants' epidemic durations was the one with yesterday's new cases as the predictor (see Fig. 2B), whose relative likelihood to outperform all the other regression models (i.e., AIC weight, Burnham & Anderson, 2002) was over 99.9%. Consistent with one might expect from the ideal observer, the more numerous yesterday's new cases, the longer participants' duration predictions (main effect $\beta = 0.075$, 95% CI [0.046, 0.104], $p < .001$). This influence was even stronger in each participant's later than earlier test days (interaction with diary number, $\beta = 0.037$, 95% CI [0.022, 0.051], $p < .001$).

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2. Human predictions are based on similarity-based function learning

How did participants then use the information of daily new cases to compute their predictions of the epidemic duration? We tested four computational models that differ in the level of theory that the observer's prediction process is based on (see Fig. 3A). That is, whether the observer infers from descriptive prior beliefs of duration distributions (the duration-prior model), uses a mechanistic internal model to simulate epidemic dynamics (the epidemiologist model), or generalizes patterns from past observations in a more theory-free way (two function learning models), and in the final case, whether the generalization is rule-based (the rule-FL model) or similarity-based (the similarity-FL model). We fit each model to participants' duration predictions for each test day using maximum likelihood estimates. AIC and cross-validated log-likelihood were used as the metrics of goodness-of-fit for model comparisons. According to both metrics, similarity-FL was the best-fitting model (see Fig. 3C and Fig. S2). The relative likelihood of similarity-FL outperforming the other three models (i.e., AIC weight) was over 99.9%.

A comparison between data and model fits (see Fig. 3B) shows how the other three models fit worse to the data. Among them, the epidemiologist model's duration predictions decreased much faster with time than participants' duration predictions did. As time elapsed, the epidemiologist model's prediction for the zero-case day effectively converged to a specific date, which deviated from the ever-postponing zero-case dates in participants' predictions (i.e., the postpone-again bias in Fig. 2A).

The major failure of the duration-prior model was its inability to capture the epidemic-related fluctuations in participants' predictions. Because the model only took into account a prior distribution of durations and the number of elapsed days since epidemic onset, its predictions would not be subject to the same influence of epidemic statistics as in the human data (see Fig. 2B).

The differences between the two function learning models (rule-RL and similarity-RL) were more subtle. Following a sharp decrease for a few days, the daily new cases entered a relatively stable period from February 19 to February 29. It was during this period that rule-RL had a remarkable under-shoot of participants' duration predictions, probably because according to the rule, the daily new cases would continue decreasing once starting to decrease. In contrast, the similarity-RL model, which had minimal assumptions about the temporal pattern of daily new cases, closely matched the human data all way through.

3.3. Psychological traits or states that influence individuals' predictions

The analyses above focused on participants' collective performance in epidemic predictions. Now, we turn to their individual differences, first screening the psychological traits or states that may correlate with individuals' duration predictions.

Among the 19 baseline measures of psychological traits or states (including subscales, see Table S1), after FDR correction for multiple comparisons, only two measures were significantly correlated with individuals' mean duration predictions (averaged across each participant's 14 test days): negative affect (DASS, Pearson's $r = .141$, 95% CI [0.046, 0.234], FDR

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

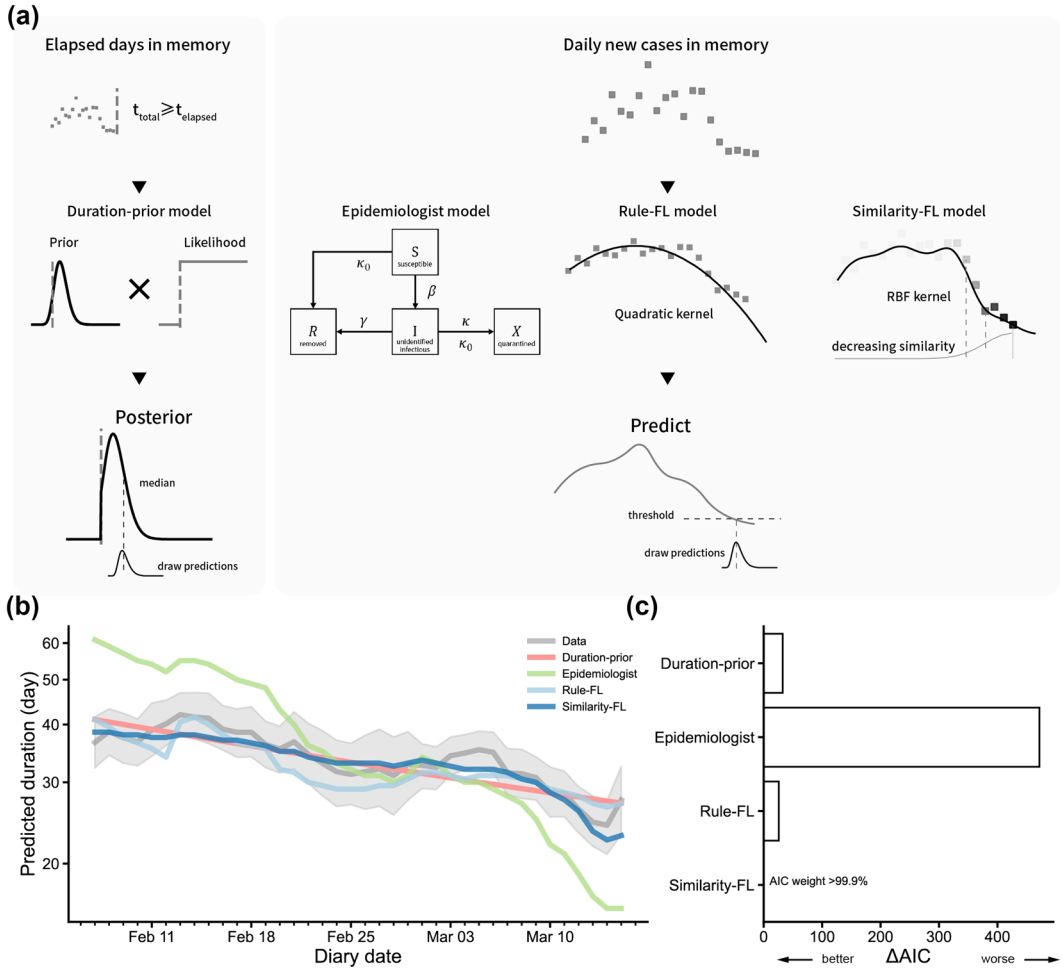

Fig. 3. Comparison of four computational models.

Note. Panel A: Illustration of the assumptions of the four models. The duration-prior model combines the number of elapsed days with a prior distribution of epidemic durations following Bayesian inference. The other three models rely on the same time series of daily new cases to predict the remaining duration of the epidemic but differ in the computations involved. The epidemiologist model assumes an observer who computes the dynamics of disease transmission similar to what epidemiologists do. The rule-FL and similarity-FL models are both observers who learn the trend of daily new cases (i.e., function learning) but differ in whether a specific functional form is assumed for the trend (rule-FL) or the generalization is simply similarity-based (similarity-FL). See text. Panel B: Data versus model fits for the duration predictions. All 418 participants were treated as a collective mind, with their duration predictions averaged separately for each diary date. Shadings for the data denote 95% CI. Panel C: Model comparison results. Smaller ΔAIC indicates better fits. Among the three models, the similarity-FL model fit best to participants' duration predictions.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

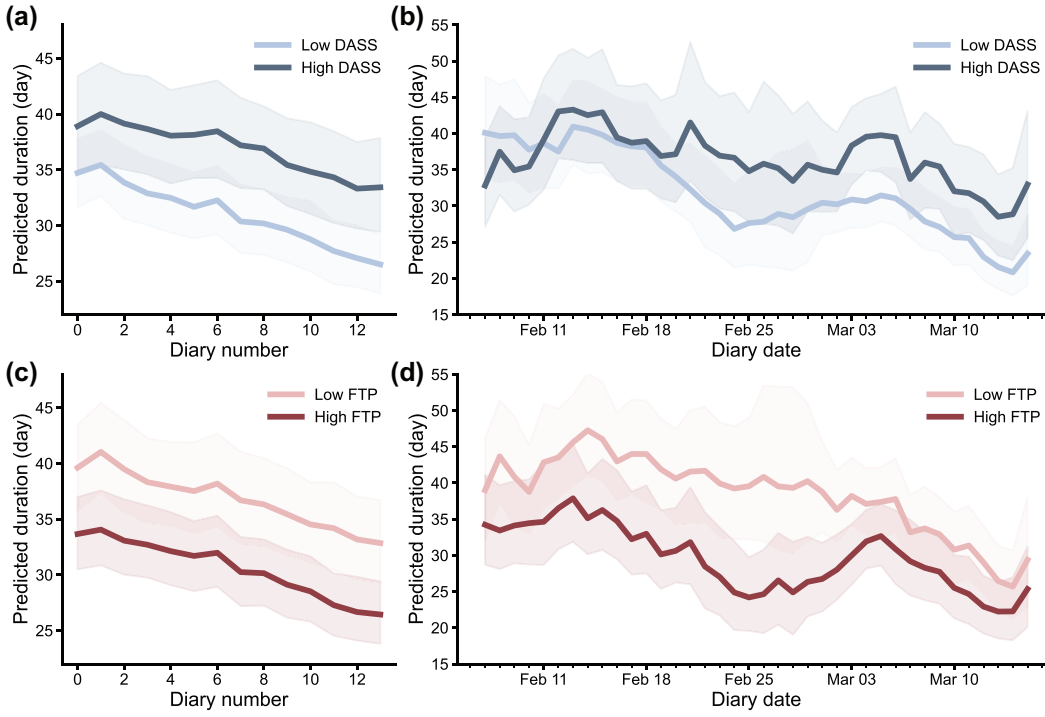

Fig. 4. Effects of negative affect and future time perspectives.

Note. Panel A, B: The duration predictions of participants with lower versus higher negative affect (lower vs. higher DASS scores). Panel C, D: The duration predictions of participants with limited versus extended future time perspectives (lower vs. higher FTP scores). Duration predictions are plotted against diary number (A, C) or diary date (B, D). Shadings denote 95% CI. See text for the statistical differences in the left panels.

corrected $p = .037$) and the perspectives of future time (FTP, Pearson's $r = -.154$, 95% CI [−0.246, −0.059], FDR corrected $p = .030$).

Next, we used linear mixed-effects regression (see Supplemental Methods) for individuals' predictions to further investigate the main effects of DASS and FTP scores as well as their interactions with daily new cases and the diary number (out of the individual's 14-day test period). The effects of daily new cases and diary numbers themselves were consistent with our results for the group data. That is, the more numerous yesterday's new cases, the longer individual participants' duration predictions ($\beta = 0.075$, 95% CI [0.044, 0.106], $p < .001$) and this influence was stronger for later diary number (interaction: $\beta = 0.035$, 95% CI [0.020, 0.051], $p < .001$). Individuals' duration predictions also decreased with diary number ($\beta = -0.069$, 95% CI [−0.094, −0.044], $p < .001$).

DASS and FTP had different effects on individuals' predictions. DASS had a significant interaction with diary number ($\beta = 0.035$, 95% CI [0.008, 0.062], $p = .011$) but a lack of significant main effect ($\beta = 0.099$, 95% CI [−0.003, 0.201], $p = .057$). As Fig. 4A shows,

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DASS scores modulated how fast the individual's duration predictions decreased with diary number: Those with higher negative affect (i.e., higher DASS scores) decreased their duration predictions more slowly with time. In contrast, FTP scores had a significant main effect ($\beta = -0.104$, 95% CI [-0.202, -0.006], $p = .037$). As Fig. 4C shows, participants with a more limited perspective for the future (i.e., lower FTP scores) predicted the epidemic to last longer. The interaction effect of DASS and the main effect of FTP were also visible in the plots of participants' duration predictions as a function of diary date, separately for participants with lower and higher DASS (see Fig. 4B) or FTP (see Fig. 4D).

3.4. Individual differences in the change of predictions over time

We have shown that psychological traits or states might influence not only an individual's mean prediction, but also how the individual's prediction changes over the 14-day test period. Next, we look further into such changes, focusing on individual differences in possible non-monotonic changes.

For each participant, we fit a quadratic function to the normalized time series of the duration predictions and used the two fitted parameters to describe the temporal pattern of the participant's duration predictions (see Supplemental Methods): trend index a_1 (increasing or decreasing), and shape index a_2 (*U*-shaped or inverted-*U*-shaped). Among the 412 participants who had at least 8 days of valid diary data, the overall trend was decreasing for 67% participants and increasing for 33% participants; the overall shape was *U*-shaped for 45% participants and inverted-*U*-shaped for 55% participants.

We computed the correlations between the two parameters and the 19 psychological traits or states measured in the baseline assessment (with FDR correction for 2×19 comparisons). Consistent with DASS's effects described earlier (see Fig. 4A), we found that the trend index a_1 had small but statistically significant positive correlations with DASS (Pearson's $r = .228$, 95% CI [0.134, 0.318], FDR corrected $p < .001$). Similar correlations were also found for two other measures of negative affect: the perceived stress (PSS) and trait anxiety (T-AI; Pearson's $r = .175$, 95% CI [0.079, 0.268], FDR corrected $p = .007$; $r = .157$, 95% CI [0.061, 0.251], FDR corrected $p = .018$, respectively). None of the correlations with the shape index a_2 passed the FDR correction for significance. In brief, consistent with our regression analysis above, the more negative an individual's affect was, the more likely the individual's duration prediction would increase with time.

3.5. The "pessimistic-me bias" in predicting epidemic duration

Apart from reporting the epidemic duration they themselves considered to be most likely, participants also estimated how most people would predict the duration. The differences between the former and the latter are shown in Fig. 5A. According to a Wilcoxon signed-rank test on each participant's difference (averaged over their 14 days test period), participants' estimation of their own "most likely" predictions was significantly longer than "most people's predictions" ($Z = 3.12$, $p = .002$). That is, participants expected themselves to be more pessimistic than others.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

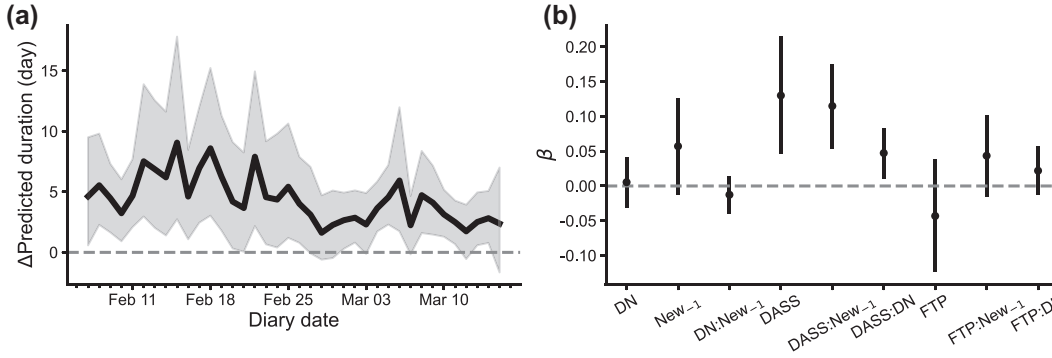

Fig. 5. The pessimistic-me bias.

Note. Panel A: The pessimistic-me bias. The difference between participants' estimation of their own "most likely" predictions and "most people's predictions" is plotted for each diary date. Black line stands for the mean prediction differences, shadings denote 95% CI. Panel B: Effects of diary number DN, yesterday's new cases New_{-1}, individual DASS and FTP on the "pessimistic me" bias. β is the standardized coefficient of the linear mixed-effects model. Error bars represent 95% CI of the coefficients. Participants with higher negative affect were aware of their pessimism, while participants with more extended future time perspectives were not aware of their optimism. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

We also tested how this pessimistic-me bias (the difference between predictions of "most likely" and "most people's") might covary with the individual's DASS or FTP scores. According to linear mixed-effects regression (see Supplemental Methods and Fig. 5B), the magnitude of the bias increased with DASS ($\beta = 0.13$, 95% CI [0.046, 0.215], $p = .003$). That is, to a greater extent did participants with higher negative affect expect their own predictions to be more pessimistic than most people. This effect of DASS was even stronger in the participant's later tests (DASS by diary number interaction: $\beta = 0.047$, 95% CI [0.011, 0.083], $p = .010$) and when yesterday's new cases were higher (DASS by yesterday's new cases interaction: $\beta = 0.115$, 95% CI [0.054, 0.176], $p < .001$). In contrast, FTP scores did not have a significant main effect ($\beta = -0.043$, 95% CI [−0.124, 0.038], $p = .297$) or any significant interactions ($p > .10$) on the magnitude of the pessimistic-me bias. Nor did the diary number ($\beta = 0.005$, 95% CI [−0.031, 0.042], $p = .772$) or yesterday's new cases ($\beta = 0.057$, 95% CI [−0.012, 0.127], $p = .108$) have significant main effects on the bias. In other words, participants with higher negative affect were aware of their pessimism, while participants with more extended future time perspectives were not aware of their optimism.

4. Control experiment 1: Prediction from minimal experience under different observations

There were two major limitations with our survey study. First, it was possible that our findings were specific to the epidemic scenario instead of more generally about prediction from minimal experience. Second, participants' observations during the epidemic outbreak

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

were uncontrollable, with the daily new cases almost consistently decreasing (though with small fluctuations) throughout the period of our survey, which could hardly constitute a strong test against coincidence. To address these two limitations, we performed a control experiment where the prediction task of our survey study was replicated with a cover story that had nothing to do with epidemics and where both the original time series of observations and a new time series with a different trend were tested.

4.1. Methods

4.1.1. Participants

There were 63 participants who completed this study. Participants were Chinese-speaking adults, recruited through the I-love-experiments platform (http://aishiyan.bnu.edu.cn/), an online participant pool maintained by Beijing Normal University. The study had been approved by the Institutional Review Boards of School of Psychological and Cognitive Sciences at Peking University. All participants provided informed consent online. Participants were compensated for their time, receiving 3.5 RMB for completing the experiment.

4.1.2. Design and procedure

Participants were told that there was an unknown object moving in an unknown space. The object started to move in the first trial and would ultimately stop. In each trial (a virtual day), participants were presented with the object's traveling distance on yesterday. During the first 16 trials (learning phase), participants only viewed the distance data and did not need to respond. After the learning phase was a 40-trial prediction phase, where participants were asked to predict the duration (i.e., number of days) needed for this object to stop. There were also two attention check questions, one in the learning phase and the other in the prediction phase. The experiment would end when participants failed in any of the attention check questions.

Participants were randomly assigned into two groups, with different time series of daily traveling distance presented to different groups. In the COVID group (34 participants), the time series was the same as the daily new cases experienced by participants in the survey study during the first wave of COVID-19 in China, though the cover story was now irrelevant to epidemics. In the plateau group (29 participants), the time series was the same as the COVID group for the first 16 trials (learning phase) but increased and decreased much more slowly afterward, as if staying at the plateau (Fig. 6A).

4.1.3. Statistical and modeling analysis

An LMM analysis was performed to examine how the last trial's observation and different time series might influence participants' predictions (see Supplement for a full description of the LMM).

Similar to the survey study, we fit the duration-prior model, the similarity-FL model, and the rule-FL model to predictions aggregated across participants, assuming that participants in different groups shared the same model parameters. We did not fit the epidemiologist model due to its inapplicability outside the epidemic scenario. We used the AIC of similarity-FL

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

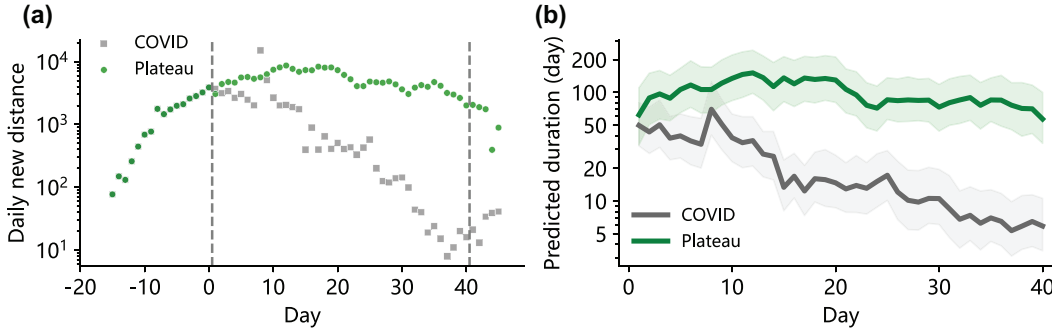

Fig. 6. The two observation conditions and participants' predictions in control experiment 1.

Note. Panel A: The sequence of daily new distances observed by each experimental group. In the learning phase, participants were sequentially presented with the object's traveling distances on 16 days. On each trial of the prediction phase, participants were informed about the object's traveling distance on the previous day (trial) and were asked to predict the number of days until the object stops (i.e., the days needed for the daily new distance to drop to 0). The observations used in the COVID group (gray square) were the same as the daily new cases in the survey study. The observations for the plateau group (green circle) agreed with those of the COVID group on the first 16 days (i.e., before day 0) but then slowly increased and decreased. Panel B: Predicted durations on each virtual day (trial). Solid lines denote the geometric mean predictions. Shadings denote 95% CI.

model as a reference and calculated each model's ΔAIC as its difference from the reference. Smaller ΔAIC indicates a better fit.

4.2. Results

As expected, the predicted durations in the COVID and the plateau groups did not differ on day 1 (t (62) = −0.57, p = .57), to which point the two groups still had the same observations. As the observed daily traveling distances in the two groups diverged, the two groups showed different prediction patterns (Fig. 6B). In both groups, yesterday's traveling distance influenced participants' predicted durations. The farther the traveling distance, the longer the predicted durations (main effect $\beta = 0.385$, 95% CI [0.154, 0.617], $p = .001$). The influence of yesterday's traveling distance was weaker on later virtual days (interaction $\beta =$ –0.094, 95% CI [−0.174, −0.015], $p = .020$). Such effects were similar for the two groups: neither the interaction between group and yesterday's traveling distance ($\beta = 0.340$, 95% CI [−0.009, 0.689], $p = .056$) nor the three-way interaction of group, yesterday's traveling distance, and virtual day number ($\beta =$ –0.101, 95% CI [−0.245, 0.044], $p = .172$) reached significance. This similar dependence on yesterday's traveling distance across groups suggests some common prediction algorithm for different observations.

Further modeling analysis replicated our finding in the survey study that the similarity-FL model best fit participants' predicted durations (Fig. 7). Note that we assumed the same parameters for the COVID and the plateau groups (i.e., a common prediction algorithm for different observations). In this way, the group differences in predicted durations could only arise from the different groups' different inputs to the model. Because the duration-prior

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

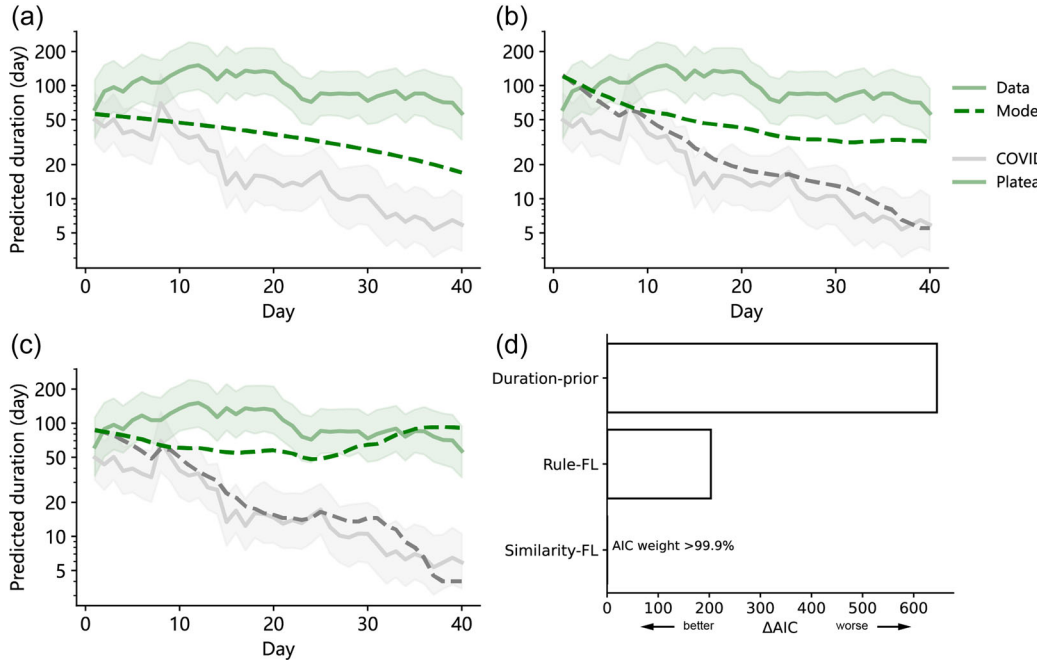

Fig. 7. Model fitting and comparison results of control experiment 1.

Note. Panels A to C: Data versus model fits of the predicted durations for the duration-prior, rule-FL, and similarity-FL models. All participants were treated as a collective mind, with their duration predictions averaged separately for each group and each day (trial). Solid lines denote data and dashed lines denote group model fits. Shadings for the data denote 95% CI. Panel D: Model comparison results. Smaller ΔAIC indicates better fits. Among the three models, the similarity-FL model fit best to participants' duration predictions.

model only used the prior distribution of durations and elapsed days to predict, it failed to capture the differences between the two groups at all (Fig. 7A). The rule-FL model predicted some group differences but had a large deviation from data for the COVID group (Fig. 7B). In contrast, the similarity-FL model well captured the prediction patterns in both groups (Fig. 7C). That is, when people have minimal experience, they use similarity-based generalization to predict the future, regardless of the task scenario or the observed time series.

5. Control experiment 2: Prediction from different prior knowledge

In the main study and control experiment 1, we found converging evidence that human prediction from minimal experience is similarity-based generalization of patterns in recent history. This cognitive algorithm is in sharp contrast with those used in predictions from rich prior knowledge, such as everyday statistics and intuitive physics. However, the prediction task we used also differed from the typical tasks in everyday statistics or intuitive physics in a few aspects. To prove that it is minimal versus rich experience instead of the particular task

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that makes the difference, we performed control experiment 2, directly testing the impact of prior knowledge in our task. Participants performed the same duration prediction task as in control experiment 1, except that they were first informed of the distribution of the historical durations. By manipulating the prior distribution as well as the time series observed by participants, we could understand how using past time series to predict the future may be influenced by prior knowledge and may differ between prediction from minimal (control experiment 1) and rich (control experiment 2) experience.

5.1. Methods

5.1.1. Participants

There were 90 participants who completed this study. Among them, one participant who entered the same numerical answer across all 40 trials was excluded, and three participants whose mean answers were outside 3 SD of the group average were excluded, resulting in 86 participants in total. Participants were Chinese-speaking adults, recruited through the Nao-Dao research platform (https://www.naodao.com/), an online participant pool. The study had been approved by the Institutional Review Boards of School of Psychological and Cognitive Sciences at Peking University. All participants provided informed consent online. Participants were compensated for their time, receiving 3.5 RMB for completing the experiment.

5.1.2. Design and procedure

Participants were randomly assigned into four conditions: two time series (COVID vs. plateau, the same as control experiment 1) by two priors (Long vs. Short duration), thus a 2×2 between-group design. The experimental procedure was the same as that of control experiment 1, except that before the learning phase participants were shown a histogram of 100 durations and were told that the histogram visualized the distribution of the number of days that similar objects traveled before stopping. The Short prior was generated with random samples from a truncated Gamma distribution, which ranged between 20 and 70 days and had a mean duration of 43 days. The Long prior was generated by adding 50 days to the samples in the Short prior (Fig. 8A). To make sure that participants had learned the prior distribution of durations, we inserted an additional attention check question about the duration prior before the learning phase.

5.2. Results

We found that the group differences initially appeared as the differences between prior conditions, with the predicted duration on day 1 being longer in the Long prior condition than the Short prior condition (main effect of prior in two-way ANOVA: F (1, 82) = 34.73, $p <$ .001). As time went on, groups with the same prior but different observations diverged, while groups with the same observations but different priors tended to converge (Fig. 8B).

In control experiment 1 where no prior information about duration was available, the influence of yesterday's traveling distance on the predicted duration decreased with time in a similar way for both the COVID and plateau groups (i.e., negative interaction between yesterday's traveling distance and virtual day number, but a lack of three-way interaction with observation

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

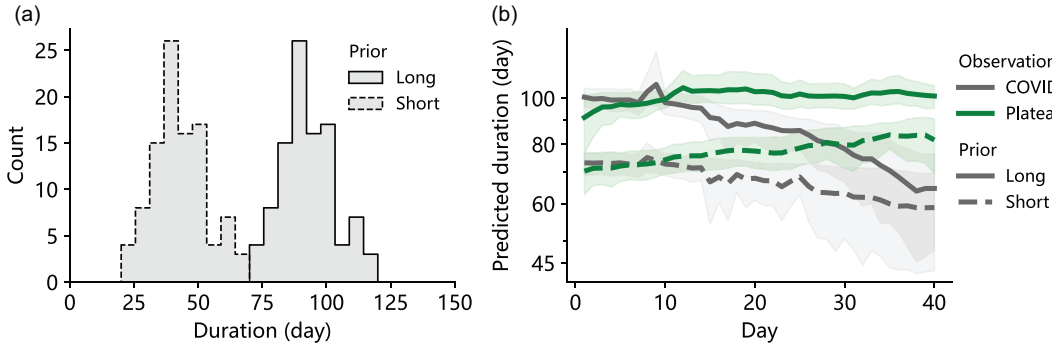

Fig. 8. The two prior conditions and participants' predictions in control experiment 2.

Note. Panel A: Two different prior distributions of durations presented to different participants. The Short prior had a mean duration of 43 days after day 0. The Long prior was generated by adding 50 days to the Short prior. Panel B: Participants' predicted durations on each day (trial), separately for different observation (gray for the COVID group and green for the plateau group) by prior (solid lines for the Long prior and dashed lines for the Short prior) conditions. Shadings denote 95% CI.

group). In contrast, here we found a four-way interaction between yesterday's traveling distance, virtual day number, the observation condition, and the prior condition ($\beta = 0.160$, 95% CI [0.047, 0.273], $p = .006$), indicating that yesterday's traveling distance influenced the predicted duration in difference temporal patterns for the four groups (2 observation conditions by 2 prior conditions). The effect of yesterday's traveling distance increased with time for the COVID with Long prior group ($\beta = 0.060$, 95% CI [0.033, 0.088], $p < .001$), decreased with time for the plateau with Long prior group ($\beta = -0.077$, 95% CI [–0.148, –0.007], $p = .003$), and was almost constant over time for the two Short prior groups (COVID with Short prior: $\beta = -0.002$, 95% CI [–0.028, 0.024], $p = .877$; plateau with Short prior: $\beta = 0.020$, 95% CI [–0.051, 0.090], $p = .582$). In other words, when equipped with prior knowledge, people may use a more complicated algorithm to predict the future, which integrates daily observations with the prior distribution of durations and whose usage of daily observations varies with the prior.

6. Discussion

During the first wave of COVID-19 outbreaks in China, people knew it was caused by a new virus and witnessed prevention measures they had never experienced or even heard of, such as the lockdown of a whole city of a 10-million population. Through a daily survey study spanning 40 days at this early phase of COVID-19 in China and two control experiments, here we show what information and what cognitive computations people use for predicting the duration of the ongoing epidemic, and more generally, for prediction from minimal experience.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6.1. Epidemic prediction as real-life function learning

In this real-life prediction task, it is impressive that people can make reasonable predictions responsive to the daily epidemic data of COVID-19, even though no information about the epidemic is presented in our survey so that people have to figure out by themselves which information in the news is relevant to the task at hand.

Our finding that people do not seem to compute epidemic dynamics but rely on a more heuristic way is in sharp contrast with what is found in intuitive physics. When predicting daily physical events, people draw predictions in a way that is similar to an "intuitive physics engine" that can simulate the future based on a mechanistic theory of physical laws (Battaglia et al., 2013; Kubricht et al., 2017). Such kind of mental simulation actually relies on one's experience in the everyday physical world (Kaiser, Jonides, & Alexander, 1986). In contrast, when facing an epidemic caused by a new virus, people have limited experience as well as insufficient knowledge of the underlying epidemic dynamics, which impedes a simulation-engine-like reasoning process. Besides, the stochastic nature of biological disease transmission is also different from the relatively deterministic physical events people experience from their daily life. The stochasticity, along with the complex interpersonal interactions in the society, makes the computations rather expensive for a mental epidemic simulation engine. That is probably why predicting epidemic dynamics may tax cognitive computations different from intuitive physics.

The lack of experience in epidemic durations also prevents people from directly inferring from a descriptive theory of duration distributions that directly combine the prior durations with elapsed time, as they do for the prediction of everyday events (Griffiths & Tenenbaum, 2006, 2011; McGuire & Kable, 2013). What people do in our task is closest to function learning that generalizes past daily case numbers to extrapolate the future. Our control experiment 1 further confirms this similarity-based generalization under minimal experience. When provided with different trends of daily observations in a different scenario irrelevant to epidemics, participants show the same prediction behavior—generalizing past observations with similarity-based function learning.

In the framework of Bayesian regression (Lucas et al., 2015; Rasmussen & Williams, 2005), different rule-based models can be implemented with Bayesian regression with different kernels. The similarity-based function learning we found is equivalent to assuming a radial-basis-function kernel, imposing minimal constraints on the temporal pattern except for smooth changes. The use of this kernel has been reported in laboratory function learning tasks (Quiroga et al., 2018; Schulz et al., 2017), but mainly as complements to more rule-based learning (e.g., linear or periodic trends). Given that people can choose suitable kernels for different everyday situations (Quiroga et al., 2018), a solo use of this minimal-constraint kernel like ours may thus be motivated by the prediction situation from minimal experience.

The winning of the similarity-based model is confined to the scenario where people have minimal experience. In situations people have rich experience with, the duration-prior model is likely to outperform the similarity-FL model. Imagine that we are watching a 100-meter race. Now the legendary world champion, Usain Bolt, is temporarily behind other runners. Despite this unfavorable observation, we might still predict that he would finish in less than 10 s and win the race, because we have rich experience about his past performance. Such

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

prediction may be better captured by the duration-prior model, where the prior belief about the durations exerts a strong influence on the prediction of future durations.

6.2. Individual differences and biases in epidemic prediction

In our real-life prediction problem, the future people need to predict is part of their lives and thus carries emotional values. Compared with individuals with less negative affect, we found that individuals with higher negative affect update their predictions with time more slowly, as if they hesitated to believe that the epidemic is approaching its end day after day. It is probably because individuals with higher negative affect weigh positive information less during their belief updating (Hein, de Fockert, & Ruiz, 2021; Korn, Sharot, Walter, Heekeren, & Dolan, 2014). This finding is also reminiscent of an influential model suggesting that exposure to stress or negative affect enhances emotional sensitivity but impairs cognitive computations critical for simulating future predictions (Liu et al., 2022; van Marle, Hermans, Qin, & Fernández, 2009). In contrast, individuals' perspectives for the future do not seem to influence the updating but the mean value of epidemic predictions: More future-oriented people predict the epidemic to end sooner. It implies that future orientation perspectives may enhance individuals' expectation for good outcomes.

Negative affect and future time perspectives also differ in their effects on the "pessimistic-me bias," the phenomenon that people consider themselves to be more pessimistic than the majority when predicting the epidemic duration. This bias is augmented by more negative affect, but unaffected by future time perspectives. It seems that individuals with more negative affect (depression, anxiety, or stress) are aware of their own biased beliefs of a more uncertain and pessimistic world, which again agrees with the enhanced emotional sensitivity but impaired cognitive computations under stressful and anxious conditions as discussed above.

6.3. Limitations and future directions

That our prediction task is situated in real life—during the outbreak of an epidemic—instead of in artificial laboratory settings enhances the ecological validity of our conclusions but meanwhile brings several methodological limitations that need to be addressed in future research. One limitation is the lack of manipulation of the epidemic trend: The daily new cases were almost consistently decreasing (though with small fluctuations) throughout the period of our survey. But this limitation has been partly addressed by our control experiments, where similarity-based generalization also holds for time series with nonmonotonic trends. A second limitation is that we could not rule out all possible forms of rule-based inference. The quadratic rule we tested was chosen to capture the intuition of a bell-shaped curve of the new cases. Besides, in situations where people have life experience about the observed time series, a combination of rule-based and similarity-based inference (implemented as Gaussian process with compound kernel functions) has been found to account well for human prediction of future trends (Quiroga et al., 2018). A third limitation of our study is about the interpretation of individual differences. In our survey study, FTP and DASS were selected post hoc from a large number of psychological scales for their correlation with individual

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

duration predictions. Whether and how these two factors really influence people's predictions awaits further investigation.

The results of the two control experiments have strengthened our conclusions about prediction from minimal experience, to which the epidemic prediction in the main study is a special case. In contrast to prediction from minimal experience, when people are provided with prior statistics of the to-be-predicted duration, they could integrate the duration prior with the observed time series to predict the future. This integration involves a potentially complicated generative model that, to our knowledge, has seldom been treated in the Bayesian literature of human cognition. Its underlying cognitive computation deserves further research.

Another future question is whether people may switch from similarity-based generalizations to more model-based predictions (as for everyday statistics or intuitive physics) after they have acquired more experience about the situation, such as after experiencing more waves of COVID-19. An analogy is the development of Theory of Mind in preschoolers, or how young children infer others' mental states (Perner & Davies, 1991). Three-year-old children, as naïve mindreaders with limited knowledge about other people's mental states, use a "copy theory" to infer others' mind, assuming that other people think exactly as themselves. As their knowledge of people's mind accumulates, till 5 years old, children form a "theory theory," a folk psychological theory about what other people are thinking. For many real-world prediction problems, adults might go through a similar process of developing theories from experience.

Acknowledgments

We thank Nigela Ahemaitijiang, Yuanqing Chang, Chao Chen, Hongmei Lin, Haining Ren, Haowen Su, Hui Wang, and Xuan Yang for helping with participant recruitment or online data collection, and Zihui Hua for APA style formatting. This work was supported by the National Natural Science Foundation of China (32171095 and 31871101 to HZ, 31871116 to LY, 32130045 to SQ, 31871121 to XZ), funding from Peking-Tsinghua Center for Life Sciences (to HZ), the Clinical Medicine Plus X - Young Scholars Project in Peking University, the Fundamental Research Funds for the Central Universities (to LY), and Beijing New Sunshine Foundation (to XZ, LY, and HZ).

Open practices statement

The study was not preregistered. The data and code are publicly accessible at https://osf.io/wu4s3/.

Note

1 We limit our attention to a narrow definition of human cognition, that is, not including more perceptual predictions.

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allen, L. J. S. (1994). Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences*, *124*(1), 83–105. https://doi.org/10.1016/0025-5564(94)90025-6

Au, T. K., Chan, C. K. K., Chan, T., Cheung, M. W. L., Ho, J. Y. S., & Ip, G. W. M. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology*, *57*(1), 1–19. https://doi.org/10.1016/j.cogpsych.2008.03.002

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18327–18332. https://doi.org/10.1073/pnas.1306572110

Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 16732–16738. https://doi.org/10.1073/pnas.2006520117

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag. https://doi.org/10.1007/b97636

Burstein, P. (2003). The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, *56*(1), 29–40. https://doi.org/10.2307/3219881

Carstensen, L. L., & Lang, F. R. (1996). *Future orientation scale* [Unpublished manuscript]. Stanford, CA: Stanford University.

Chu, L. C., & Kao, H. S. (2005). The moderation of meditation experience and emotional intelligence on the relationship between perceived stress and negative mental health. *Chinese Journal of Psychology*, *47*(2), 178–194.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, *369*(6500), eabb9789. https://doi.org/10.1126/science.abb9789

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(4), 968–986. https://doi.org/10.1037/0278-7393.23.4.968

Estrada, E. (2020). COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. *Physics Reports*, *869*, 1–51. https://doi.org/10.1016/j.physrep.2020.07.005

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Fu, Han, Haowei Wang, Xiaoyue Xi, Adhiratha Boonyasiri, Yuanrong Wang, Wes Hinsley, Keith J. Fraser, Ruth McCabe, Daniela Olivera Mesa, Janetta Skarp, Alice Ledda, Tamsin Dewé, Amy Dighe, Peter Winskill, Sabine L. van Elsland, Kylie E. C. Ainslie, Marc Baguelin, Samir Bhatt, Olivia Boyd, Nicholas F. Brazeau, Lorenzo Cattarino, Giovanni Charles, Helen Coupland, Zulma M. Cucunuba, Gina Cuomo-Dannenburg, Christl A. Donnelly, Ilaria Dorigatti, Oliver D. Eales, Richard G. FitzJohn, Seth Flaxman, Katy A. M. Gaythorpe, Azra C. Ghani, William D. Green, Arran Hamlet, Katharina Hauck, David J. Haw, Benjamin Jeffrey, Daniel J. Laydon, John A. Lees, Thomas Mellan, Swapnil Mishra, Gemma Nedjati-Gilani, Pierre Nouvellet, Lucy Okell, Kris V. Parag, Manon Ragonnet-Cronin, Steven Riley, Nora Schmit, Hayley A. Thompson, H. Juliette T. Unwin, Robert Verity, Michaela A. C. Vollmer, Erik Volz, Patrick G. T. Walker, Caroline E. Walters, Oliver J. Watson, Charles Whittaker, Lilith K. Whittles, Natsuko Imai, Sangeeta Bhatia, and Neil M. Ferguson (2021). Database of epidemic trends and control measures during the first wave of COVID-19 in mainland China. *International Journal of Infectious Diseases*, *102*, 463–471. https://doi.org/10.1016/j.ijid.2020.10.075

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Griffiths, T. L., & Tenenbaum, J. B. (2011). Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*, *140*(4), 725–743. https://doi.org/10.1037/a0024899

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition*, *157*, 61–76. https://doi.org/10.1016/j.cognition.2016.08.012

Hein, T. P., de Fockert, J., & Ruiz, M. H. (2021). State anxiety biases estimates of uncertainty and impairs reward learning in volatile environments. *NeuroImage*, *224*, 117424. https://doi.org/10.1016/j.neuroimage.2020.117424

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*, 262–267. https://doi.org/10.1038/s41586-020-2404-8

Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, *14*(4), 308–312. https://doi.org/10.3758/bf03202508

Kalichman, S. C., & Cain, D. (2005). Perceptions of local HIV/AIDS prevalence and risks for HIV/AIDS and other sexually transmitted infections: Preliminary study of intuitive epidemiology. *Annals of Behavioral Medicine*, *29*(2), 100–105. https://doi.org/10.1207/s15324796abm2902_4

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, *44*(3), 579–592. https://doi.org/10.1017/S0033291713001074

Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, *21*(10), 749–759. https://doi.org/10.1016/j.tics.2017.06.002

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development: Coexistence of natural and supernatural explanations. *Child Development*, *83*(3), 779–793. https://doi.org/10.1111/j.1467-8624.2012.01743.x

Li, J., Lai, S., Gao, G. F., & Shi, W. (2021). The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature*, *600*(7889), 408–418. https://doi.org/10.1038/s41586-021-04188-6

Liu, L., Wu, J., Geng, H., Liu, C., Luo, Y., Luo, J., & Qin, S. (2022). Long-term stress and trait anxiety affect brain network balance in dynamic cognitive computations. *Cerebral Cortex*. *32*(14), 2957–2971. https://doi.org/10.1093/cercor/bhab393

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*(5), 1193–1215. https://doi.org/10.3758/s13423-015-0808-5

Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, *368*(6492), 742–746. https://doi.org/10.1126/science.abb4557

McGuire, J. T., & Kable, J. W. (2013). Rational temporal predictions can underlie apparent failures to delay gratification. *Psychological Review*, *120*(2), 395–410. https://doi.org/10.1037/a0031910

Merow, C., & Urban, M. C. (2020). Seasonality and uncertainty in global COVID-19 growth rates. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(44), 27456–27464. https://doi.org/10.1073/pnas.2008590117

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, *32*(7), 1133–1147. https://doi.org/10.1080/03640210802353016

Perner, J., & Davies, G. (1991). Understanding the mind as an active information processor: Do young children have a "copy theory of mind"? *Cognition*, *39*(1), 51–69. https://doi.org/10.1016/0010-0277(91)90059-D

Quiroga, F., Schulz, E., Speekenbrink, M., & Harvey, N. (2018). Structured priors in human forecasting. *BioRxiv*, 285668. https://doi.org/10.1101/285668

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian processes for machine learning*. MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

Rubin, G. J., Amlot, R., Page, L., & Wessely, S. (2009). Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: Cross sectional telephone survey. *BMJ*, *339*, b2651 https://doi.org/10.1136/bmj.b2651

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive Psychology*, *99*, 44–79. https://doi.org/10.1016/j.cogpsych.2017.11.002

Sigelman, C. K., & Glaser, S. E. (2019). Characterizing children's intuitive theories of disease: The case of flu. *Cognitive Development*, *52*, 100809. https://doi.org/10.1016/j.cogdev.2019.100809

Smith, K. A., Battaglia, P. W., & Vul, E. (2018). Different physical intuitions exist between tasks, not domains. *Computational Brain & Behavior*, *1*(2), 101–118. https://doi.org/10.1007/s42113-018-0007-3

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y1–Y2)*. Palo Alto, CA: Consulting Psychologists Press.

Stojic, H., Schulz, E., Analytis, P. P., & Speekenbrink, M. (2020). It's new, but is it good? How generalization and uncertainty guide the exploration of novel options. *Journal of Experimental Psychology: General*, 1878–1907. https://doi.org/10.1037/xge0000749

Taouk, M., Lovibond, P. F., & Laube, R. (2001). *Psychometric properties of a Chinese version of the 21-item Depression Anxiety Stress Scales (DASS21)*. Sydney, NSW: Transcultural Mental Health Centre. Cumberland Hospital.

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From specificity to sensitivity: How acute stress affects amygdala processing of biologically salient stimuli. *Biological Psychiatry*, *66*(7), 649–655. https://doi.org/10.1016/j.biopsych.2009.05.014

Wang, X., Wang, X., & Ma, H. (1999). *Handbook of Mental Health Assessment Scales (revised)*. Retrieved from http://ir.bjmu.edu.cn/handle/400002259/91600

Zheng, X., Shu, L., Zhang, A., Huang, G., Zhao, J., Sun, M., Fu, Y., Li, H., & Xu, D. (1993). Report of the state-trait anxiety in Changchun. *Chinese Mental Health Journal*, *02*, 60–62.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1
Figure S2
Figure S3
Figure S4
Supplemental Online Material

15516709, 2023, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.13294 by Peking University Health, Wiley Online Library on [15/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License